The Laws of Robotics

Isaac Asimov

It isn’t easy to think about computers without wondering if they will ever “take over.”

Will they replace us, make us obsolete, and get rid of us the way we got rid of spears and tinderboxes?

If we imagine computerlike brains inside the metal imitations of human beings that we call robots, the fear is even more direct. Robots look so much like human beings that their very appearance may give them rebellious ideas.

This problem faced the world of science fiction in the 19208 and 19308, and many were the cautionary tales written of robots that were built and then turned on their creators and destroyed them.

When I was a young man I grew tired of that caution, for it seemed to me that a robot was a machine and that human beings were constantly building machines. Since all machines are dangerous, one way or another, human beings built safeguards into them.

In 1939, therefore, I began to write a series of stories in which robots were presented sympathetically, as machines that were carefully designed to perform given tasks, with ample safeguards built into them to make them benign.

In a story I wrote in October 1941, I finally presented the safeguards in the specific form of “The Three Laws of Robotics. “ (I invented the word robotics, which had never been used before.)

Here they are:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where those orders would conflict with the First Law.

3. A robot must protect its own existence except where such protection would conflict with the First and Second Law.

These laws were programmed into the computerized brain of the robot, and the numerous stories I wrote about robots took them into account. Indeed, these laws proved so popular with the readers and made so much sense that other science fiction writers began to use them (without ever quoting them directly—only I may do that), and all the old stories of robots destroying their creators died out.

Ah, but that’s science fiction. What about the work really being done now on computers and on artificial intelligence? When machines are built that begin to have an intelligence of their own, will something like the Three Laws of Robotics be built into them?

Of course they will, assuming the computer designers have the least bit of intelligence. What’s more, the safeguards will not merely be like the Three Laws of Robotics; they will be the Three Laws of Robotics.

I did not realize, at the time I constructed those laws, that humanity has been using them since the dawn of time. Just think of them as “The Three Laws of Tools,” and this is the way they would read:

1. A tool must be safe to use.

(Obviously! Knives have handles and swords have hilts. Any tool that is sure to harm the user, provided the user is aware, will never be used routinely whatever its other qualifications.)

2. A tool must perform its function, provided it does so safely.

3. A tool must remain intact during use unless its destruction is required for safety or unless its destruction is part of its function.

No one ever cites these Three Laws of Tools because they are taken for granted by everyone. Each law, were it quoted, would be sure to be greeted by a chorus of “Well, of course!”

Compare the Three Laws of Tools, then, with the Three Laws of Robotics, law by law, and you will see that they correspond exactly. And why not, since the robot or, if you will, the computer, is a human tool?

But are safeguards sufficient? Consider the effort that is put into making the automobile safe—yet automobiles still kill 50,000 Americans a year. Consider the effort that is put into making banks secure—yet there are still bank robberies in a steady drumroll. Consider the effort that is put into making computer programs secure—yet there is the growing danger of computer fraud.

Computers, however, if they get intelligent enough to “take over,” may also be intelligent enough no longer to require the Three Laws. They may, of their own benevolence, take care of us and guard us from harm.

Some of you may argue, though, that we’re not children and that it would destroy the very essence of our humanity to be guarded.

Really? Look at the world today and the world in the past and ask yourself if we’re not children—and destructive children at that—and if we don’t need to be guarded in our own interest.

If we demand to be treated as adults, shouldn’t we act like adults? And when do we intend to start?